



Article

DispHred: A Server to Predict pH-Dependent Order–Disorder Transitions in Intrinsically Disordered Proteins

Jaime Santos [†] , Valentín Iglesias [†] , Carlos Pintado , Juan Santos-Suárez 
and Salvador Ventura ^{*} 

Institut de Biotecnologia i Biomedicina and Departament de Bioquímica i Biologia Molecular,
Universitat Autònoma de Barcelona, 08193 Barcelona, Spain; jaime.santos@uab.es (J.S.);
valentin.iglesias@uab.cat (V.I.); carlos.pintado@e-campus.uab.cat (C.P.); juansans98@gmail.com (J.S.-S.)

^{*} Correspondence: salvador.ventura@uab.es; Tel.: +34-93-586-8956

[†] Authors contributed equally.

Received: 25 July 2020; Accepted: 11 August 2020; Published: 13 August 2020



Abstract: The natively unfolded nature of intrinsically disordered proteins (IDPs) relies on several physicochemical principles, of which the balance between a low sequence hydrophobicity and a high net charge appears to be critical. Under this premise, it is well-known that disordered proteins populate a defined region of the charge–hydropathy (C–H) space and that a linear boundary condition is sufficient to distinguish between folded and disordered proteins, an approach widely applied for the prediction of protein disorder. Nevertheless, it is evident that the C–H relation of a protein is not unalterable but can be modulated by factors extrinsic to its sequence. Here, we applied a C–H-based analysis to develop a computational approach that evaluates sequence disorder as a function of pH, assuming that both protein net charge and hydrophobicity are dependent on pH solution. On that basis, we developed DispHred, the first pH-dependent predictor of protein disorder. Despite its simplicity, DispHred displays very high accuracy in identifying pH-induced order/disorder protein transitions. DispHred might be useful for diverse applications, from the analysis of conditionally disordered segments to the synthetic design of disorder tags for biotechnological applications. Importantly, since many disorder predictors use hydrophobicity as an input, the here developed framework can be implemented in other state-of-the-art algorithms.

Keywords: intrinsically disordered proteins; pH; bioinformatics; disorder prediction; conditional folding; machine learning

1. Introduction

Intrinsically disordered proteins (IDPs) are a class of polypeptides that do not require a defined folded structure to execute their biological activities [1–3]. The plasticity of these biomolecules allows them to interact with structurally diverse partners, and they are often involved in the wiring of protein networks, acting both as central hubs and as molecular switches [4]. The unfolded nature of IDPs is intrinsically encoded in their primary sequence, which is generally enriched in ionizable and polar residues and depleted of hydrophobic amino acids [5]. Thus, IDPs' extended conformation depends both on electrostatic repulsions between uncompensated charges and on a low hydrophobicity, which prevents extensive protein compaction [6].

Based on the balance between attractive and repulsive forces in IDPs, Uversky and coworkers proposed that they populated a distinct region in the mean net charge–hydropathy (C–H) phase space diagram and demonstrated that an empirical boundary line was enough to discriminate folded and disordered proteins [6]. Under that premise, the disordered nature of a polypeptide sequence can

be predicted by evaluating its C–H relationship in the aforementioned attraction–repulsion scheme. The C–H plot analysis has been applied for disorder prediction; it lies behind the popular FoldIndex algorithm [7], and it is also computed by other multiparametric software [8].

More than 50 predictors, based on different molecular principles, have been developed to assess protein disorder, thus providing a robust toolbox for identifying natively unfolded proteins or their regions [8–10]. Besides, new tools able to reverse-engineer the above-mentioned principles into a sequence that now allows for the artificial design of disordered protein segments [11,12]. Nevertheless, most of these methods are blind to the protein context, even if IDPs are extremely sensitive to environmental fluctuations [13,14]. Ligands, binding partners, or solvent conditions such as ions concentration or pH, have been reported to induce conditional folding in IDPs [6,15,16]. Therefore, it is surprising to find out that those effects have been mostly disregarded in state-of-the-art computational approaches. Indeed, it is immediately evident that the C–H relationship of a given protein is not constant since both protein net charge and hydrophobicity can be modulated by factors that are extrinsic to the sequence.

In recent work, we showed that the solution's pH effect on IDPs solubility is not restricted to its effect on the charge of ionizable residues since the pH also modulates the sequence hydrophobicity [17], a traditionally neglected effect. Driven by this simple idea, we revisit here the C–H concept, on the evidence that both protein net charge and hydrophobicity are dependent on pH. By delineating a boundary condition similar to the one described by Uversky [6], we demonstrate that IDPs' pH-induced folding can be predicted just by evaluating the pH dependence of the C–H space diagram. This allowed us to develop DispHred, a first computational approach to predict protein disorder as a function of pH. DispHred is freely available for academic users at <https://ppmclab.pythonanywhere.com/DispHred>. We hope that the data we present here may prompt the development of a new generation of disorder predictors that include solvent conditions on their pipelines.

2. Results

2.1. Validation of a pH-Dependent Hydrophathy Scale for C–H Plot-Based Predictions

The original C–H analysis was developed using the Kyte–Doolittle hydrophathy scale to calculate the mean hydrophobicity of protein sequences [6,7,18]. Here, we implement a novel amino acid pH-dependent hydrophathy scale developed by Zamora and co-workers [19], based on implicit solvation calculations, that allow us to evaluate the effect of the solution pH on sequence hydrophobicity. As a first step in developing our approach, we needed to assess the performance of this pH-dependent scale for C–H plot-based order–disorder predictions at neutral pH. Uversky and Dunker did an extensive analysis of 19 diverse hydrophathy scales to compare their performance in C–H plot-based predictions [20]. They reported that the Guy hydrophathy scale [21] had the highest discriminative power, while Kyte–Doolittle performance was in the average of the 19 scales. Additionally, they developed a new scale that provided the best order–disorder discrimination (IDP–Hydrophathy) [20].

We compared the pH-dependent hydrophathy (pH-dependent) scale with the Kyte–Doolittle, Guy, and IDP–Hydrophathy scales. First, we normalized the four scales between 0 and 1, assigning a value of 1 to the highest hydrophobicity. Then, we calculated the values for the pH-dependent scale at pH 7.0. We found the highest correlation with the Guy scale ($R^2 = 0.72$), followed by the Kyte–Doolittle ($R^2 = 0.60$) and the IDP–hydrophathy ($R^2 = 0.51$) scales (Figure 1A–C). The correlation between the Guy and Kyte–Doolittle scales is $R^2 = 0.78$. In contrast, as it happens for the pH-dependent scale, the correlation between the IDP–hydrophathy and the Guy or the Kyte–Doolittle scales is low, with $R^2 = 0.52$ and $R^2 = 0.33$, respectively. These low correlations stem mostly from the fact that, counter-intuitively, the IDP–hydrophathy scale considers Pro as the most hydrophilic residue, with a value of 0 in our normalized scale. Removing Pro from the correlation between the pH-dependent and IDP–hydrophathy scales increases R^2 to 0.70 and arbitrarily assigning this residue a value of 0 in the pH-dependent scale (pH–Pro-corrected scale) results in an $R^2 = 0.74$ (Supplementary Figure S1A).

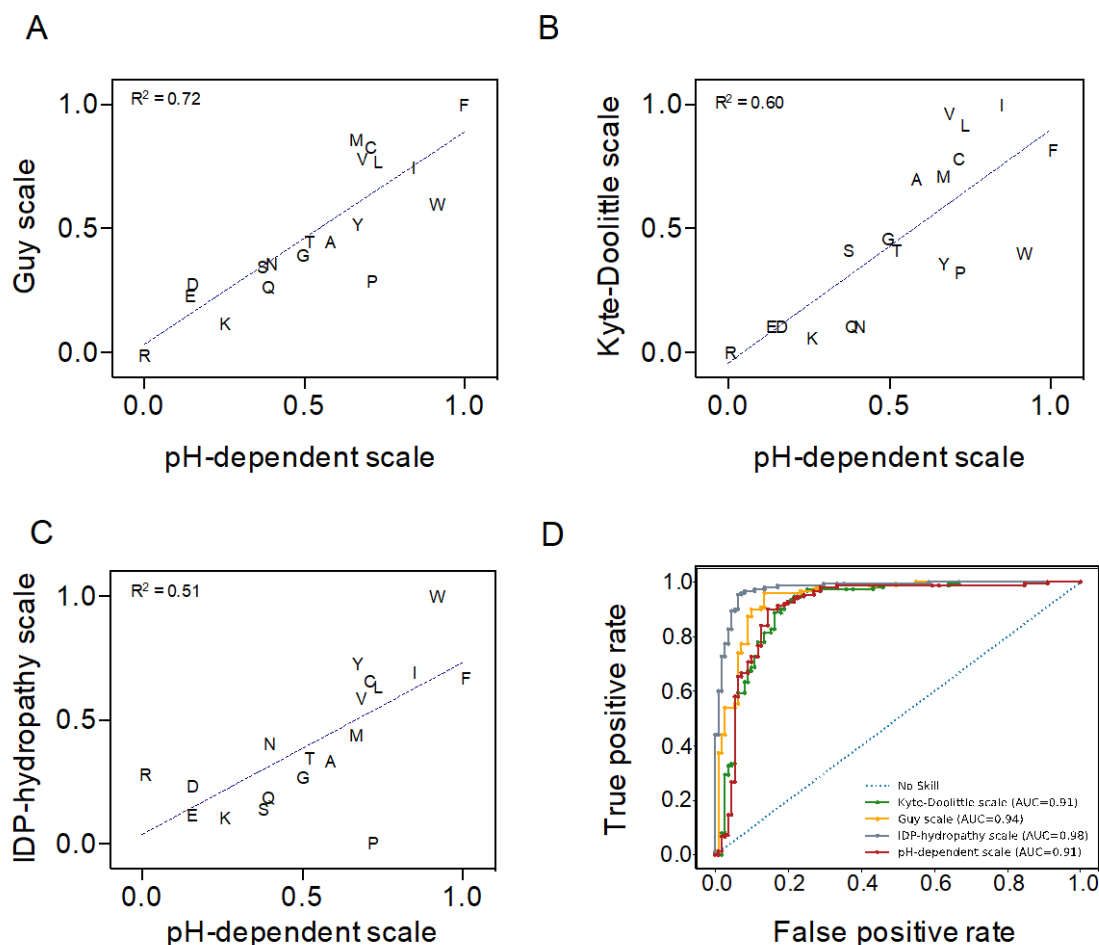


Figure 1. Comparison of four different hydropathy scales at pH 7.0. Correlation between pH-dependent scale and (A) Guy, (B) Kyte–Doolittle, and (C) IDP–hydropathy scales. Letters correspond to the amino acid one-letter code. Hydropathy values are normalized between 0 and 1, corresponding to the minimum and maximum values for each scale. The R^2 value of the linear regression is shown in each graph. (D) Receiver Operating Characteristic curves showing the performance of the four scales in discriminating a dataset of fully disordered ($n = 111$) and single-chain folded ($n = 150$) proteins.

We next ensembled a dataset of 111 experimentally validated fully disordered proteins and 150 folded single-chain proteins with X-ray resolved structures (Supplementary Table S1) to test the discriminatory power of the four scales in a C–H plot analysis. The ability to classify ordered and disordered sequences of each scale was assessed by applying a Receiver Operating Characteristic (ROC) method. The associated area under the curve (AUC) was used as a sensitivity–specificity reporter. The pH-dependent and the Kyte–Doolittle scales showed an identical discriminatory potential ($AUC = 0.91$), while the Guy and IDP–hydropathy scales demonstrated slightly higher performances ($AUC = 0.94$ and 0.98 , respectively) (Figure 1D). The pH–Pro-corrected scale exhibited an $AUC = 0.95$ (Supplementary Figure S1B), which suggests that the minimal value assigned to Pro in the IDP–hydropathy scale contributes to its higher discrimination.

Overall, the analysis suggested that the pH-dependent scale compared well with the other analyzed scales at pH 7.0, with a discriminatory power identical to the widely employed Kyte–Doolittle scale. Thus, this scale will allow us to extend the C–H predictive potential to the full pH scale without compromising the performance at neutral pH significantly. Despite its higher discrimination, we preferred to not use the pH–Pro-corrected scale and keep the hydropathy value obtained from implicit solvation calculations for Pro residues [19].

2.2. C–H Space Phase Diagram and Order–Disorder Boundary Condition Can Anticipate pH-Induced Order–Disorder Transition of IDPs

Next, we explored whether the C–H model would be a reliable tool to predict the pH-dependent order–disorder transition in IDPs. To that end, we performed a bibliographic search of structural data on IDPs that suffer a conditional folding at specific pHs. We collected 59 bibliographic pH datapoints for 7 disordered proteins and peptides (Figure 2 and Supplementary Figure S2). For each point, we calculated the protein net charge per residue (NCPR) and protein mean hydrophobicity at the given pH $\langle H_{pH} \rangle$. To do so, NCPR is calculated using the Henderson–Hasselbach equation, and $\langle H_{pH} \rangle$ is computed according to the pH-dependent scale developed by Zamora and co-workers [19]. We plotted each datapoint in a 3-axis scatter plot according to its pH, $\langle H_{pH} \rangle$ and NCPR, with the dot's color indicating whether the protein was folded or disordered in this condition (Figure 2).

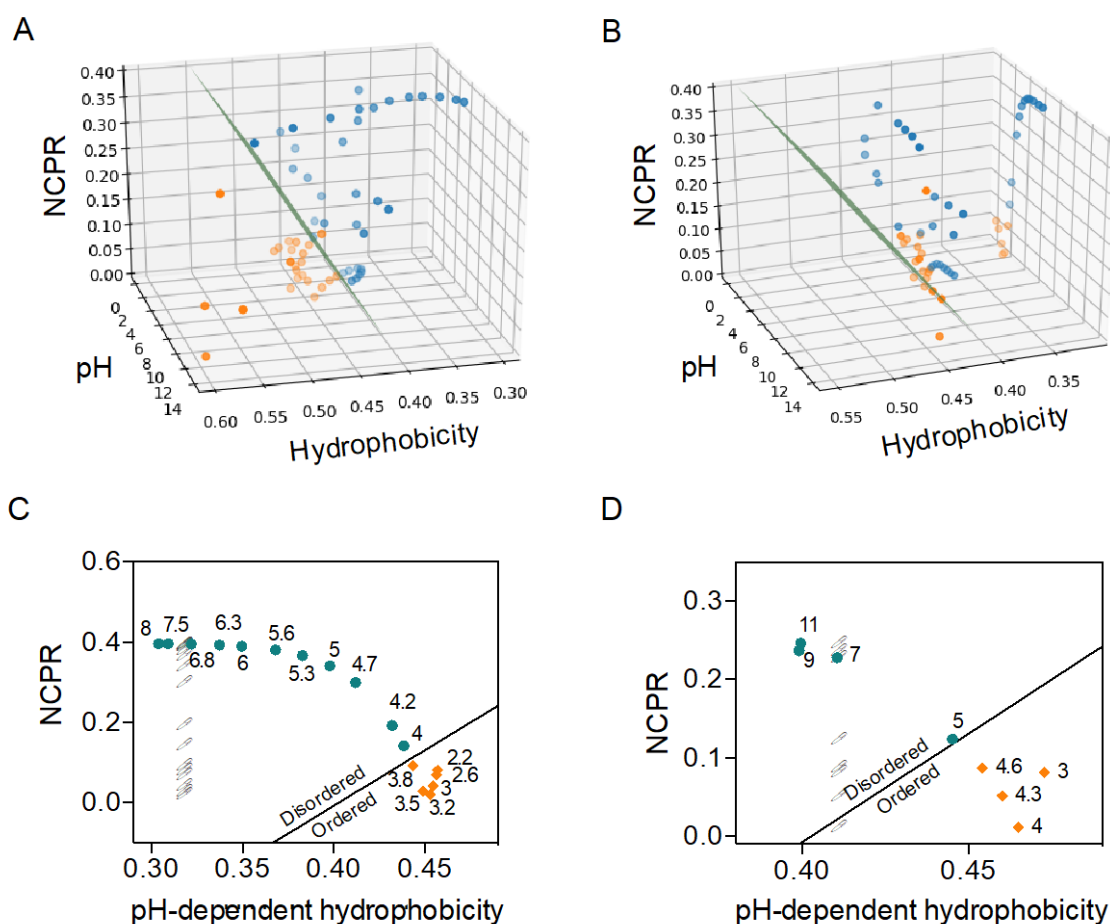


Figure 2. Charge–Hydropathy-based analysis of pH modulated order–disorder transitions. Three-dimensional C–H plots containing 59 datapoints of 7 proteins at different pHs, computing pH influence over (A) sequence net charge per residue (NCPR) and hydrophobicity, or (B) assuming constant hydrophobicity values (as calculated at pH 7.0). Blue and orange points correspond to conditions in which protein/peptides are disordered and folded, respectively. The green surfaces delimit the boundary conditions between folded and disordered proteins defined by Equation (1). (C,D) Two-dimensional C–H plots of (C) prothymosin and (D) PEST-c-myc using the same color pattern than in panels A and B for folded–unfolded datapoints. A solid line represents the boundary condition. Open circles represent the same data points assuming constant hydrophobicity values (as calculated at pH 7.0).

To develop a consistent C–H-based order–disorder classification for the experimental data, we sought to seek the order–disorder boundary condition that allowed the maximal separation between the two states. Since the datasets for the different proteins diverged in size, nature, and source,

we assumed that a classic iterative analysis might lead to overfitting and/or result in a biased boundary condition in case some data points were misclassified.

To minimize such limitations, we applied a support vector machine (SVM) learning strategy, a supervised feedforward network specifically designed to build a binary classifier and retrieve the boundary condition that maximizes the separation between observations [22,23]. SVM-based analysis reduces overfitting and tolerates a certain degree of misclassified data points without forcing a bias, being robust classification strategies, and increasing their predictive potential when applied to new observations, especially near the boundary condition. Additionally, since SVM analysis takes into account a slight uncertainty and misclassification, it also provides a margin near the boundary line (dashed lines in Supplementary Figure S3) that can be used as a confidence interval in a subsequent classification of new data points in predictive applications.

By using the above-described SVM-based analysis, we identified a linear boundary condition defined by Equation (1)

$$Dis_{pH} = 2.775 < H_{pH} > - |NCPR| - 1.118 \quad (1)$$

that successfully discriminates between folded and disordered proteins with a Matthews Correlation Coefficient of 0.97 (Supplementary Figure S3A, Table 1). Note that our boundary condition is reasonably similar to that previously defined by Uversky and co-workers (Equation (2)) [6] for order–disorder classifications at neutral pH.

$$I = 2.785 < H > - | < R > | - 1.151 \quad (2)$$

Table 1. Performance of pH-dependent and pH-independent hydrophobicity approaches in predicting pH-conditioned order–disorder transitions in a C–H analysis by applying Equation (1). Unfolded sequences correctly predicted to be unfolded were classified as true positives. The highest values for each measure are indicated in bold.

Measure	pH-Dependent Hydrophobicity	pH-Independent Hydrophobicity
Sensitivity	1.00	1.00
Specificity	0.96	0.21
Precision	0.97	0.65
False Discovery rate	0.03	0.35
Accuracy	0.98	0.68
F1 Score	0.99	0.79
Matthews Correlation Coefficient	0.97	0.37

$\langle H \rangle$ and $\langle R \rangle$ correspond to the mean hydrophobicity and mean charge at neutral pH, respectively.

In contrast, applying the same SVM analysis but considering that hydrophobicity is independent of pH, we did not observe a consistent classification of the datapoints—Matthews Correlation Coefficient of 0.6—and the boundary line didn't satisfy the C–H relationship (Supplementary Figure S3B, Supplementary Table S2).

As shown in Figure 2A, the boundary plane defined by Equation (1) satisfactorily delimited folding–unfolding transitions for the analyzed IDPs, with only one datapoint wrongly predicted and still reasonably close to the boundary. This translates into 98% accuracy in predicting the proteins' conformational states at any given pH (Table 1). On the contrary, by considering that hydrophobicity is independent of pH (and computing its value at pH 7.0 and under the same boundary condition as Equation (1)), we observed that the NCPR change alone could not discriminate between folded and disordered sequences (Figure 2B and Table 1). This observation evidences the importance of modeling the pH-dependent hydrophobicity when predicting protein disorder.

Prothymosin is a classic example of an IDP at neutral pH that experiences a conditional folding at low pHs, characterized by the gain of α -helical structure [24]. The transition occurs between pH 3.5 and pH 5.0, with prothymosin being fully folded below pH 3.5 and fully unfolded above pH 5.0. In a two-dimensional projection of the data points for this protein, we can observe that all folded

points fall below the boundary line, being thus accurately predicted (Figure 2C). We also observed that our pH-dependent C–H representation also succeeds in delineating the transition range (pH 3.5–5). Similarly, the disordered PEST region (201–268) from human c-Myc oncoprotein collapses into a folded conformation at pHs below 4.8 [25], a transition that is successfully identified by our pH-dependent C–H ratio (Figure 2D). Note that the same analysis considering a constant hydrophobicity is blind to these structural conversions (open circles in Figure 2C,D). The same trend can be observed in the two-dimensional C–H plots of the other five protein sets in Figure 2A,B (Supplementary Figure S2).

These data demonstrate that the effect of pH on IDPs' conditional folding can be successfully predicted by applying a pH-dependent C–H analysis. With these results in hand, we aimed to develop a computational tool for predicting protein disorder that considers implicitly the solution pH, which we named DispHred.

2.3. Rational and Implementation of DispHed, a pH-dependent Predictor of Sequence Disorder

DispHred uses the C–H space diagram analysis proposed by Uversky and co-workers and later implemented in FoldIndex [6,7]. Nevertheless, instead of considering constant net charges and hydrophobicity for each analyzed sequence, DispHred assumes that the solution pH modulates both parameters. Thus, DispHred computes the protein NCPR and the mean hydrophobicity of a sequence as a function of pH. Then, DispHred applies the boundary condition defined by Equation (1) to separate folded and disordered proteins. Dis_{pH} positive values correspond to sequences classified as folded and negative values to those classified as disordered at the analyzed pH or pH range. The SVM approach provides a margin of ± 0.02 around the boundary line used as a confidence interval in the classification.

DispHred calculates the Dis_{pH} score for all the analyzed pHs, profiling the pH-dependence disorder of a protein sequence, and thus including the pH dimension in the classical C–H phase diagram. DispHred runs a user-defined sliding window that enables the analysis of the folded/disordered regions in a protein sequence at every requested pH. Sequence stretches fall in three classes: i) regions that are predicted to be always folded in the analyzed pH interval, ii) regions that are predicted to be always disordered in this pH interval, and iii) regions whose folded/disordered conformation is modulated by the pH.

DispHred is available at <https://ppmclab.pythonanywhere.com/DispHred>. DispHred is free for academic users and does not require login. In the input page the user can (i) introduce a sequence in FASTA format or insert a valid Uniprot Accession number, (ii) select the pH range and step size for the analysis or type a single specific pH and (iii) select the sliding window size (Figure 3A). After running the program, the user will be redirected to a results page containing the report of the analysis (Figure 3B): Dis_{pH} scores, mean hydrophobicity, and NCPR for each of the analyzed pHs, a graph showing Dis_{pH} score as a function of pH, and clickable links that redirect to the sequence profile prediction at each desired pH. The protein regions exhibiting pH-dependent and pH-independent folded/disordered conformations are colored on top of the input sequence.

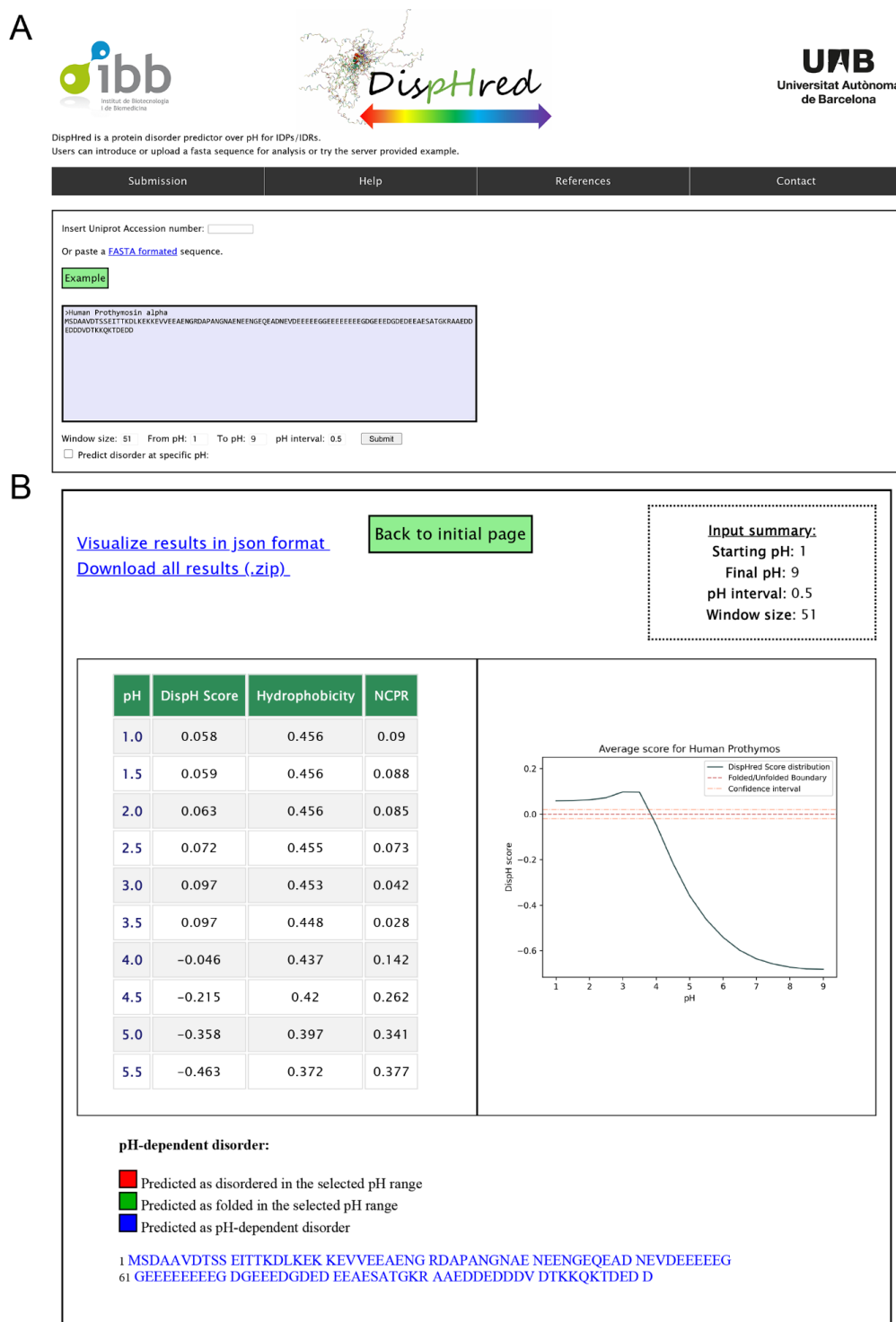


Figure 3. DispHred web server interface. **(A)** Web input page. The user can paste their sequence or insert a valid Uniprot Accession number. DispHred works by default by checking disorder in a range of pHs but allows users to test values at a specific pH. By default, a 51-residue sliding window is populated, but users can personalize its length. **(B)** Web results page for a selected range of pHs. Two clickable links appear on the upper left part of the screen with a JSON file or a ZIP file containing DispHred calculations and generated figures. On the lower left part, the table shows the DispHred, hydrophobicity, and NCPR average scores for each pH. Clicking each pH will open a figure representing the Dis_{pH} score variation along the sequence for the selected pH. On the lower right part, a figure representing the Dis_{pH} average score for each pH is shown. Scores above the red dashed line indicate predicted order. On the bottom of the screen, folded, disordered, and conditionally disordered regions are indicated in the sequence in green, red and blue, respectively.

Users can retrieve all data in a JavaScript Object Notation (JSON) file or download all the generated data in a compressed ZIP file. A clickable example is provided in the input page to illustrate DispHred outputs.

3. Discussion

Structural disorder is a fundamental trait of protein biology that complements the activities of structured proteins and domains by contributing flexibility and plasticity [26–28]. In contrast to folded proteins, IDPs exist as ensembles sampling a wide range of dynamic conformations in which the bulk of the primary sequence is highly exposed to the solvent. Accordingly, IDPs' properties display little dependence on structural elements and can be inferred from the primary sequence, which has allowed the design of computational tools for predicting, designing, and analyzing protein disorder [8,9,11]. At the same time, IDPs are extremely sensitive to environmental conditions; an effect often disregarded in predictive approaches. Among the different parameters that may affect IDPs' properties, the solution pH has a significant impact, mainly due to the high prevalence of ionizable residues in these polypeptides [14,15,17,29]. In this work, we demonstrated that the effect of pH on the disordered nature of a protein sequence could be easily predicted by evaluating the changes in protein charge and hydrophobicity as a function of this parameter. Even if the effect of pH over net charge is well-recognized, hydrophobicity is usually considered to be a pH-independent property. However, we found that the evaluation of the pH-dependent hydrophobicity is fundamental for the accuracy of the order/disorder prediction in any given condition.

The analysis of the local or global hydrophobicity of protein sequences is a pivotal stage in many *in silico* pipelines aimed to predict protein disorder and its associated properties. A significant number of disorder predictors, such as FoldIndex or PONDR, rely on the direct or indirect analysis of hydrophobicity, a property that is also used to predict folding upon binding, RNA–, DNA-interactions or post-translational modification sites in IDPs [7–9,30–34]. Thus, the identification of hydropathy scales suitable for such analyses attracted significant attention in the past [20]. Our results indicate that by applying a recently developed pH-dependent hydropathy scale, the contribution of this predictive physicochemical property to disorder prediction can be extended to the full pH scale. Thus, the implementation of pH-dependent hydropathy scales, like the one we use here, may increase applicability in currently available algorithms.

pH, ion concentrations, redox state, or post-translational modifications are known regulators of protein function by controlling the switch between the disordered and folded or partially folded states of polypeptides. Thus, although the conditional disorder's prediction is a challenging task, it is fundamental to elucidate the functionality of IDPs. [13,35]. To advance in this direction, we developed DispHred, an online web server that exploits the C–H space analysis to predict protein disorder as a function of pH. Its main application is the profiling of protein disorder across a continuous pH interval, demonstrating a high accuracy in classifying the pH-modulated order–disorder transitions for sequentially unrelated model proteins and peptides. Additionally, DispHred allows the assessment of the specific protein regions contributing the most to conditional disorder.

In essence, DispHred is the first disorder predictor dedicated to evaluating the effect of the solution pH and constitutes a proof-of-concept for the implementation of this kind of approach in future predictive endeavors. Intrinsically disorder tags are increasingly used to solubilize proteins and to engineer the pharmacological properties of protein and peptide pharmaceuticals [36]. We envision that DispHred can be of significant help in these biotechnological tasks.

4. Materials and Methods

4.1. Data Collection

The dataset of 111 experimentally verified fully disordered proteins was obtained from the Disprot database (DisProt 2020_06) [37] by selecting proteins with a 100% disorder coverage. The set of 150

fully folded sequences was randomly extracted from the Protein Data Bank (PDB) under the query single-chain structures larger than 100 residues and determined by X-ray crystallography.

Data regarding the effect of pH on protein disorder was extracted from the bibliography. Data regarding the pH-dependent folding of prothymosin was obtained from the characterization of Uversky and coworkers [24]. Order–disorder pH-transition of the PEST region (201–268) from human c-Myc oncoprotein was analyzed in Mohd. Ziauddin Ansari and Rajaram Swaminathan’s study [25]. LL-37 pH-dependent helix formation was reported by Johansson and coworkers [38]. Victor Muñoz and Luis Serrano reported the effect of solution pH on a model peptide Ac-AKAAKAKAAKAKA-NH₂ [39]. Data on the pH-modulated collapse of human histones were extracted from Munishkina and coworkers [40]. The analysis of the disordered A-domain of the Toc132 receptor disorder was performed by Lynn GL Richardson, Masoud Jelokhani-Niaraki, and Matthew D Smith [41]. The conformational fluctuations of the 36-loop region of the influenza hemagglutinin were analyzed by Chavela M. Carr and Peter S. Kim [42].

4.2. DispHred: Evaluation of Hydrophobicity and Charge as a Function of pH

To analyze the lipophilicity of protein sequences, we employed the pH-dependent lipophilicity scale developed by Zamora and coworkers. They used continuum solvation calculations, which allow us to calculate the hydrophobicity of a given residue at the desired pH [19]. Then, DispHred uses a sliding window with a user-defined length to calculate the average hydrophobicity in the window and assigns it to the residue in the center. In the analysis performed in this article, we used a fixed window of 7 residues. The results were averaged to calculate the mean hydrophobicity of the sequence at the analyzed pH.

Protein NCPR is calculated by applying the Henderson–Hasselbalch equation to derive the partial charge of each ionizable residue at the analyzed pH. Then, global NCPR is calculated as the sum of all partial charges divided by the protein length. To calculate the Dis_{pH} score of a given window, the NCPR is calculated using the residues included in this particular window and its length.

4.3. Hydropathy Scales Performance Analysis at Neutral pH

We delineated a C–H plot for each of the analyzed hydropathy scales. Each scale was normalized from 0 to 1 according to the increased hydrophobicity of the protein residues; for the pH-dependent scale, we employed the values calculated at pH 7.0 [19]. The performance of the different scales was evaluated using a ROC analysis, in which the true-positive rate is plotted against the false-positive rate. The ROC analysis was performed against a dataset of 111 fully disordered proteins and 150 single-chain folded proteins. The AUC was taken as an indicator of sensitivity and sensibility.

4.4. Support Vector Machine Analysis

SVM was applied to define the optimal boundary line delimitating two classes of samples as folded or disordered. NCPR and pH-dependent hydrophobicity were calculated as previously stated for the 59 data points. Experimental data was labeled as ordered or disordered as described in the literature and employed for the machine learning process. To perform the analysis, we used the freely available machine learning library scikit-learn [43]. The SVM kernel was set to “linear” to map the data on a two-dimensional space.

4.5. DispHred: Prediction of Sequence Disorder

DispHred uses a C–H plot analysis to discriminate between folded and disordered sequences at the analyzed pH by applying a defined boundary condition. For each pH, the mean hydrophobicity ($\langle H_{pH} \rangle$) and the absolute value of the NCPR are calculated. Then, the Dis_{pH} score is obtained by applying the SVM–derived boundary condition (Equation (1)). Positive and negative values are classified as folded or disordered, respectively. DispHred calculates the Dis_{pH} score at all the pHs in the desired range to profile sequence disorder as a function of pH. DispHred also analyzes the Dis_{pH} score of the sliding windows to identify specific stretches whose disorder is affected by pH.

4.6. Performance Analysis

The sensitivity, specificity, precision, accuracy and false discovery rate when predicting order–disorder transitions was evaluated as follows: Sensitivity = $TP/(TP + FN)$; Specificity = $TN/(TN + FP)$; Precision = $TP/(TP + FP)$; Accuracy = $(TP + TN)/(TP + TN + FP + FN)$; and False Discovery Rate = $FP/(FP + TP)$. F1 Score and Matthews Correlation Coefficient were calculated as previously described in [44]. TP, TN, FP and FN correspond to true positives, true negatives, false positives and false negatives, respectively

4.7. DispHred Web-Server

The DispHred web server interface was built in HTML/CSS/JS. It uses the Django 3.0 framework working with python 3.7. The figures are generated using matplotlib library [45]. The server is platform-independent, free and open for academic users and does not require a previous login.

Supplementary Materials: Supplementary materials can be found at <http://www.mdpi.com/1422-0067/21/16/5814/s1>.

Author Contributions: Conceptualization, S.V., J.S. and V.I.; software, V.I., J.S., C.P., J.S.-S.; validation, V.I. and J.S.; data curation, V.I. and J.S.; writing—original draft preparation, V.I., J.S. and S.V.; supervision, S.V.; funding acquisition, S.V. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Spanish Ministry of Economy and Competitiveness BIO2016-78310-R to S.V. and by ICREA, ICREA-Academia 2015 to S.V., J.S. was supported by the Spanish Ministry of Science and Innovation via a doctoral grant (FPU17/01157).

Acknowledgments: We thank Jordi Pujols for insightful discussion.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

AUC	Area under the curve
C–H	Charge–hydropathy
IDP	Intrinsically disordered protein
NCPR	Net charge per residue
ROC	Receiver Operating Characteristic
SVM	Support Vector Machine
<H>	Mean hydrophobicity

References

1. Dunker, A.K.; Obradovic, Z. The protein trinity—linking function and disorder. *Nat. Biotechnol.* **2001**, *19*, 805–806. [CrossRef] [PubMed]
2. Kulkarni, V.; Kulkarni, P. Intrinsically disordered proteins and phenotypic switching: Implications in cancer. *Prog. Mol. Biol. Transl. Sci.* **2019**, *166*, 63–84. [PubMed]
3. Chen, J.; Kriwacki, R.W. Intrinsically Disordered Proteins: Structure, Function and Therapeutics. *J. Mol. Biol.* **2018**, *430*, 2275–2277. [CrossRef]
4. Wright, P.E.; Dyson, H.J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 18–29. [CrossRef]
5. Dyson, H.J. Making Sense of Intrinsically Disordered Proteins. *Biophys. J.* **2016**, *110*, 1013–1016. [CrossRef]
6. Uversky, V.N.; Gillespie, J.R.; Fink, A.L. Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins* **2000**, *41*, 415–427. [CrossRef]
7. Prilusky, J.; Felder, C.E.; Zeev-Ben-Mordehai, T.; Rydberg, E.H.; Man, O.; Beckmann, J.S.; Silman, I.; Sussman, J.L. FoldIndex: A simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* **2005**, *21*, 3435–3438. [CrossRef] [PubMed]
8. He, B.; Wang, K.; Liu, Y.; Xue, B.; Uversky, V.N.; Dunker, A.K. Predicting intrinsic disorder in proteins: An overview. *Cell Res.* **2009**, *19*, 929–949. [CrossRef] [PubMed]

9. Lieutaud, P.; Ferron, F.; Uversky, A.V.; Kurgan, L.; Uversky, V.N.; Longhi, S. How disordered is my protein and what is its disorder for? A guide through the “dark side” of the protein universe. *Intrinsically Disord. Proteins* **2016**, *4*, e1259708. [[CrossRef](#)] [[PubMed](#)]
10. Dosztanyi, Z. Prediction of protein disorder based on IUPred. *Protein Sci.* **2018**, *27*, 331–340. [[CrossRef](#)]
11. Schramm, A.; Lieutaud, P.; Gianni, S.; Longhi, S.; Bignon, C. InSiDDe: A Server for Designing Artificial Disordered Proteins. *Int. J. Mol. Sci.* **2017**, *19*, 91. [[CrossRef](#)] [[PubMed](#)]
12. Harmon, T.S.; Crabtree, M.D.; Shammass, S.L.; Posey, A.E.; Clarke, J.; Pappu, R.V. GADIS: Algorithm for designing sequences to achieve target secondary structure profiles of intrinsically disordered proteins. *Protein Eng. Des. Sel.* **2016**, *29*, 339–346. [[CrossRef](#)] [[PubMed](#)]
13. Jakob, U.; Kriwacki, R.; Uversky, V.N. Conditionally and transiently disordered proteins: Awakening cryptic disorder to regulate protein function. *Chem. Rev.* **2014**, *114*, 6779–6805. [[CrossRef](#)] [[PubMed](#)]
14. Uversky, V.N. Intrinsically disordered proteins and their environment: Effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding. *Protein J.* **2009**, *28*, 305–325. [[CrossRef](#)] [[PubMed](#)]
15. Smith, M.D.; Jelokhani-Niaraki, M. pH-induced changes in intrinsically disordered proteins. *Methods Mol. Biol.* **2012**, *896*, 223–231. [[PubMed](#)]
16. Fonin, A.V.; Stepanenko, O.V.; Sitdikova, A.K.; Antifeeva, I.A.; Kostyleva, E.I.; Polyanchko, A.M.; Karasev, M.M.; Silonov, S.A.; Povarova, O.I.; Kuznetsova, I.M.; et al. Folding of poly-amino acids and intrinsically disordered proteins in overcrowded milieu induced by pH change. *Int. J. Biol. Macromol.* **2019**, *125*, 244–255. [[CrossRef](#)]
17. Santos, J.; Iglesias, V.; Santos-Suarez, J.; Mangiagalli, M.; Brocca, S.; Pallares, I.; Ventura, S. pH-Dependent Aggregation in Intrinsically Disordered Proteins Is Determined by Charge and Lipophilicity. *Cells* **2020**, *9*, 145. [[CrossRef](#)]
18. Kyte, J.; Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, *157*, 105–132. [[CrossRef](#)]
19. Zamora, W.J.; Campanera, J.M.; Luque, F.J. Development of a Structure-Based, pH-Dependent Lipophilicity Scale of Amino Acids from Continuum Solvation Calculations. *J. Phys. Chem. Lett.* **2019**, *10*, 883–889. [[CrossRef](#)]
20. Huang, F.; Oldfield, C.J.; Xue, B.; Hsu, W.L.; Meng, J.; Liu, X.; Shen, L.; Romero, P.; Uversky, V.N.; Dunker, A. Improving protein order-disorder classification using charge-hydropathy plots. *BMC Bioinform.* **2014**, *15* (Suppl. 17), S4. [[CrossRef](#)]
21. Guy, H.R. Amino acid side-chain partition energies and distribution of residues in soluble proteins. *Biophys. J.* **1985**, *47*, 61–70. [[CrossRef](#)]
22. Vapnik, V. *Statistical Learning Theory*; Wiley: Hoboken, NJ, USA, 1998; Volume 1, p. 768.
23. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer Science & Business Media: Berlin, Germany, 2013.
24. Uversky, V.N.; Gillespie, J.R.; Millett, I.S.; Khodyakova, A.V.; Vasiliev, A.M.; Chernovskaya, T.V.; Vasilenko, R.N.; Kozlovskaya, G.D.; Dolgikh, D.A.; Fink, A.L.; et al. Natively unfolded human prothymosin alpha adopts partially folded collapsed conformation at acidic pH. *Biochemistry* **1999**, *38*, 15009–15016. [[CrossRef](#)] [[PubMed](#)]
25. Ansari, M.Z.; Swaminathan, R. Structure and dynamics at N- and C-terminal regions of intrinsically disordered human c-Myc PEST degron reveal a pH-induced transition. *Proteins* **2020**, *88*, 889–909. [[CrossRef](#)] [[PubMed](#)]
26. Oldfield, C.J.; Dunker, A.K. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu. Rev. Biochem.* **2014**, *83*, 553–584. [[CrossRef](#)] [[PubMed](#)]
27. Tompa, P. Intrinsically disordered proteins: A 10-year recap. *Trends Biochem. Sci.* **2012**, *37*, 509–516. [[CrossRef](#)] [[PubMed](#)]
28. Babu, M.M.; Van der Lee, R.; de Groot, N.S.; Gsponer, J. Intrinsically disordered proteins: Regulation and disease. *Curr. Opin. Struct. Biol.* **2011**, *21*, 432–440. [[CrossRef](#)]
29. Payliss, B.J.; Vogel, J.; Mittermaier, A.K. Side chain electrostatic interactions and pH-dependent expansion of the intrinsically disordered, highly acidic carboxyl-terminus of gamma-tubulin. *Protein Sci.* **2019**, *28*, 1095–1105. [[CrossRef](#)]
30. Xue, B.; Dunbrack, R.L.; Williams, R.W.; Dunker, A.K.; Uversky, V.N. PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochim. Biophys. Acta* **2010**, *1804*, 996–1010. [[CrossRef](#)]

31. Garner, E.; Romero, P.; Dunker, A.K.; Brown, C.; Obradovic, Z. Predicting Binding Regions within Disordered Proteins. *Genome Inform. Ser. Workshop Genome Inform.* **1999**, *10*, 41–50.
32. Iakoucheva, L.M.; Radivojac, P.; Brown, C.J.; O'Connor, T.R.; Sikes, J.G.; Obradovic, Z.; Dunker, A.K. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* **2004**, *32*, 1037–1049. [[CrossRef](#)]
33. Meng, F.; Uversky, V.N.; Kurgan, L. Comprehensive review of methods for prediction of intrinsic disorder and its molecular functions. *Cell. Mol. Life Sci.* **2017**, *74*, 3069–3090. [[CrossRef](#)] [[PubMed](#)]
34. Ward, J.J.; McGuffin, L.J.; Bryson, K.; Buxton, B.F.; Jones, D.T. The DISOPRED server for the prediction of protein disorder. *Bioinformatics* **2004**, *20*, 2138–2139. [[CrossRef](#)] [[PubMed](#)]
35. Bardwell, J.C.; Jakob, U. Conditional disorder in chaperone action. *Trends Biochem. Sci.* **2012**, *37*, 517–525. [[CrossRef](#)] [[PubMed](#)]
36. Minde, D.P.; Halff, E.F.; Tans, S. Designing disorder: Tales of the unexpected tails. *Intrinsically Disord. Proteins* **2013**, *1*, e26790. [[CrossRef](#)] [[PubMed](#)]
37. Hatos, A.; Hajdu-Soltesz, B.; Monzon, A.M.; Palopoli, N.; Alvarez, L.; Aykac-Fas, B.; Bassot, C.; Benitez, G.I.; Bevilacqua, M.; Chasapi, A.; et al. DisProt: Intrinsic protein disorder annotation in 2020. *Nucleic Acids Res.* **2020**, *48*, D269–D276. [[CrossRef](#)]
38. Johansson, J.; Gudmundsson, G.H.; Rottenberg, M.E.; Berndt, K.D.; Agerberth, B. Conformation-dependent antibacterial activity of the naturally occurring human peptide LL-37. *J. Biol. Chem.* **1998**, *273*, 3718–3724. [[CrossRef](#)] [[PubMed](#)]
39. Munoz, V.; Serrano, L. Elucidating the folding problem of helical peptides using empirical parameters. III. Temperature and pH dependence. *J. Mol. Biol.* **1995**, *245*, 297–308. [[CrossRef](#)]
40. Munishkina, L.A.; Fink, A.L.; Uversky, V.N. Conformational prerequisites for formation of amyloid fibrils from histones. *J. Mol. Biol.* **2004**, *342*, 1305–1324. [[CrossRef](#)]
41. Richardson, L.G.; Jelokhani-Niaraki, M.; Smith, M.D. The acidic domains of the Toc159 chloroplast preprotein receptor family are intrinsically disordered protein domains. *BMC Biochem.* **2009**, *10*, 35. [[CrossRef](#)]
42. Carr, C.M.; Kim, P.S. A spring-loaded mechanism for the conformational change of influenza hemagglutinin. *Cell* **1993**, *73*, 823–832. [[CrossRef](#)]
43. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
44. Chicco, D.; Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genom.* **2020**, *21*, 6. [[CrossRef](#)] [[PubMed](#)]
45. Hunter, J. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).